DOCUMENT RESUME

ED 196 206 EC 131 333

AUTHOR Foster, Carol
TITLE Relationships Among Traditional and Behavior Assessment Procedures in a Behavioral Research Program.
INSTITUTION Kansas Research Inst. for the Early Childhood Education of the Handicapped, Lawrence.
SPONS AGENCY Bureau of Education for the Handicapped (DHEW/OE), Washington, D.C.
REPORT NO ECI-112
PUB DATE 78
GRANT 300-77-0308
NOTE 24p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Disabilities; Early Childhood Education; *Evaluation Methods; *Handicap Identification; *Research Design
IDENTIFIERS *At Risk (for Handicap)

ABSTRACT

The paper is designed to lay the foundation for a traditional assessment program in a systematic research effort on the identification of and intervention with at risk and handicapped infants and young children. Initial sections review literature on the qualities of traditional psychometric assessment and on the qualities of a behavioral assessment program. It is explained in a third section that the proposed program will involve a longitudinal approach to the development of systematic and accurate procedures for identifying children and infants who are at risk for handicapping conditions, to the measurement of environmental events in home and school situations for the purpose of identifying those variables related to handicapping conditions, and to the development of and validation of intervention strategies which alleviate handicapping conditions in these children. A final section considers the relationship of traditional assessment to the behavioral research program. (SBH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED196206

RELATIONSHIPS AMONG TRADITIONAL AND BEHAVIOR ASSESSMENT PROCEDURES IN A BEHAVIORAL RESEARCH PROGRAM

by

Carol Foster

Kansas Research Institute for the Early Childhood
Education of the Handicapped

(ECI Document No. 112)

1978

Preparation of this paper was supported by a grant from the Bureau for the Education of the Handicapped (USOE 300-77-0308).

No portion of this paper may be reprinted without permission from the author

EC 131333

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Rogers-Warren

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

RELATIONSHIPS AMONG TRADITIONAL AND BEHAVIOR ASSESSMENT PROCEDURES IN A BEHAVIORAL RESEARCH PROGRAM

Carol Foster

The Research Institute for the Early Childhood Education of the Handicapped is undertaking a systematic research effort on the identification of and intervention with at-risk and handicapped infants and young children. An emphasis is to be placed on behavioral assessment procedures. With that in mind, a clear statement of the distinctions and similarities between behavioral and traditional assessment is necessary. A plan can then be designed for a behavioral assessment program, which is supplemented with a traditional assessment program. This position paper is intended to lay the fundation for a traditional assessment program in the Institute's behavioral research program.

Qualities of Traditional Psychometric Assessment

Cronbach (1970) defines a test as ". . . a systematic procedure for observing a person's behavior and describing it with the aid of a numerical scale or a category system" (p. 26). This definition covers measurement procedures as diverse as vision tests and final exams in college.

Psychometric tests make two assumptions: 1) If something exists, it exists in some amount; and 2) If something exists in some amount, it can be measured (Cronbach, p. 26).

In other words, any personal characteristic or attribute can be evaluated and translated into numerical terms. This contrasts with an impressionistic assessment that seeks a total integrated representation of an individual, frequently through clinical interpretations. There is little emphasis on translating that into a formal scoring procedure.

Psychometric testing is characterized by methodological components. One of these components is the use of formal scoring procedures and the translation of performance into a numerical score. This emphasis also allows a second major methodology of psychometric tests--the emphasis on critical validation. Ideally, this provides a potential user with information on how much error there is in the measurement and to what extent a prediction made with that instrument is likely to come true (Cronbach, 1970, p. 34).

Psychometric assessment, then, is a procedure for measuring characteristics of individuals in numerical terms so that persons using the test know how accurate their measurements are.

According to Livingston (1977, p. 314) traditional, norm-referenced psychometric tests make five major assumptions. First, a test measures some underlying variable. Second, its purpose is to reveal differences among individuals. Third, what a test measures is indicated by the size and direction of its correlations with other tests. Fourth, unless otherwise noted, individual scores are normally distributed over the range of that test. Finally, a score on a test is important only in comparison with other scores. This is usually depicted by a set of norms.

These assumptions are consistent with the description of psychometric testing given by Cronbach (1970) summarized above. Essentially, some variable is hypothesized. A method of measurement is devised. To determine if this method actually measures the variable in question, correlations are made between it and other measures of that variable. If no other quantitative measure exists, then a correlation is made between the new measure and clinical judgments (e.g., Leland, Nihira, Foster, Shellhaas, & Kagin, Note 1, pp. 21-22). The next step would be the acquisition of normative information concerning the general population on the measures developed through a sampling procedure. From these data, norms would be generated, reliability information obtained, measurement error would be ascertained, and a manual written. At this point, a measurement device is ready to evaluate the degree to which an individual possesses the hypothesized variable in relation to other members of that population.

These assumptions and procedures are essentially the methodology (in brief) used in the construction of most tests of intelligence or child development. Early measures of children's intellectual, perceptual, and psycholinguistic abilities have not empirically been found to be strongly related to the children's performance level in later years (Evans & Nelson, 1977; Ysseldyke, 1973). While it cannot be denied that early measures of child development are correlated to later measures, the relationship is tenuous, and prediction on an individual basis is not warranted. Consequently, it is almost impossible to make decisions concerning the necessity of an intervention program with a given child

unless that child scores at the extreme ends of the distribution of scores on that test. Typically, for these children, decisions concerning intervention can be made without assessment devices.

While it is true that some of the problems in predicting from these traditional assessment devices may result from limitations in the validity, stability, and reliability of the instruments (Hammill & Larson, 1974; Hammill & Wiederbolt, 1973; Sedlak & Weener, 1973), it is also possible that using assessment procedures that emphasize description of entire populations ranging from individuals with few skills to those with abundant skills will of necessity result in an instrument which does not make fine distinctions. That is, the 'at-risk' child and the 'normal' child may be within a few points of each other, and the standard error of measurement is typically larger than this span. While it is possible to have a test comprehensive enough to cover the full range of abilities and precise enough to make fine distinctions at this level, this kind of test does not currently exist, and is unlikely to be developed because it would necessitate an extremely large test, much of which would be irrelevant for any given child (Foster, Note 2). Consequently, measurement error and differences between children are difficult to distinguish and traditional assessment procedures for the identification of 'at-risk' children have very limited success.

Simeonsson (1977, pp. 30-32) summarizes the utility of psychometric assessment procedures. The advantages are that they are standardized, widely-used, and they reduce subjectivity. The disadvantages he presents

are that (a) the standardization procedures typically excluded handicapped children; (b) the norm tables do not extend to the low ranges; (c) there is no way to adapt the tests for specific handicapping conditions; and (d) the test may not depict the total capabilities of the child.

Qualities of a Behavioral Assessment Program

Traditional psychometric assessment is not adequate for researchers or educators who have had a major interest in behavior change of specific individuals rather than groups. Cronbach (1970) acknowledges that the experimental analysis of behavior has developed measurement systems similar to psychometric methods in that they involve an objective tabulation of responses. However, these methods deviate from traditional psychometric methods because the results are rarely summarized into a few scores, the measures are taken over a fairly extensive period of time, and remediation procedures are frequently included during the measurement period. He concludes that the assessment methods of experimental analysis of behavior do not strictly constitute a 'test' (Cronbach, 1970, pp. 33-34).

Behavioral assessment stems from the experimental analysis of behavior, and concentrates on the objective measurement of behavior occurrence or behavior change in the area of applied behavior analysis. While various authors have conceptualized the purposes of behavioral assessment somewhat differently, they have generally addressed the

same issues. Goldfried and Linehan (1977, pp. 16-17) describe the three purposes of behavioral assessment in a clinical setting as:

1. To determine the variables which are maintaining maladaptive behavior and the characteristics affecting the therapy procedures.
2. To select individuals for one type of treatment versus another. (This is an area of which has received little attention so far, but has potential for behavioral assessment.)
3. To evaluate the effectiveness of the treatment procedures, and to contribute to the research literature in the field.

Ciminero (1977, p. 6) also concentrated on a clinical treatment model, and was in agreement on the purposes.

Another view is provided by behavioral educators Tillman, Bersoff and Dolly (1976) who perceive testing as consisting of systematic methods for assessing student behavior. The purposes of assessment consist of obtaining a description of the behavior in question, obtaining information on the variables that affect that behavior, and determining means for changing it (p. 293). Repeated assessment over time can then be used to determine if the behavior in question has been affected by the intervention. Sulzer-Azaroff and Mayer (1977, p. 49) identify the purpose of precise behavioral measurement as a procedure for determining specifically and objectively whether applied behavior analysis programs are effective and for generating scientific predictions for the future.

Another educator, Livingston (1977), discusses behavioral assessment in terms of criterion-referenced testing. Its purposes are to estimate an individual's skills or information in absolute terms (i.e., against

some specific criterion) and to classify an individual as having achieved a specific objective or not achieved a specific objective.

While there are some different purposes for behavioral assessment, depending on whether they are more for clinical or educational assessment, there are some basic points of agreement. First, they provide a behavioral and objective description of some behavior rather than indirectly measuring some underlying variable. Second, there is an emphasis on how an individual is described or is affected by some intervention procedure, rather than on determining the statistical distribution of individuals along some measure. And third, there is an emphasis on measurement over time to identify the effectiveness of certain procedures rather than on obtaining a single measure at one time for many individuals.

Once the purposes of behavioral assessment have been established, there are many related characteristics and assumptions that distinguish it from psychometric or traditional assessment. Livingston (1977) has summarized the contrasts between psychometric tests and one form of behavioral assessment--the criterion-referenced test. His summary is given in Table I. Another column has been added to show the differences between behavioral assessment generally and psychometric tests. As can be seen from this table, the two types of assessment differ drastically. The major similarity is that they are both interested in measuring, but what they measure and how the results are used are quite different.

One major difference is that behavioral assessment is a direct measurement of the behavior in question and this behavior is a function of environmental events. The psychometric tests are indirect measures

Table I. Assumptions of Traditional and Criterion-Referenced Approaches to Educational Measurement
(Take from Livingston in Cone and Hawkins)

		Traditional	Criterion-Referenced	Behavioral Assessment
1.	A test is. . . .	An indicator of some underlying variable.	A sample of the behaviors of interest.	A sample of the behaviors of interest.
2.	The purpose of a test is. . .	To reveal differences between individuals.	To estimate a person's level of skill or achievement in absolute terms; or To classify a person as a master or nonmaster of a specified behavior.	To describe the behavior(s) of an individual; To have a basis of evaluating the effectiveness of a behavior change procedure; To aid in the selection of a treatment or educational strategy.
3.	What a test measures is indicated by. . .	Its pattern of correlations with other tests.	The specification of the domain from which items are selected; or The extent to which it classifies persons in the same way as other classification procedures.	The definitions of the behaviors measured.
4.	The distribution over persons of a variable underlying a test score. . .	Can be assumed to be normal unless there is a specific reason to believe otherwise.	Is not of interest; or Is a two-point distribution; or Cannot be assumed to have particular form.	Is of no interest; or Cannot be assumed to have any particular form.
5.	The score of a person on a test derives its meaning from. . .	Comparisons with the scores of other persons.	Its interpretation as a direct estimate of the person's absolute level of proficiency; or a conditional probability distribution of some related variable; or Comparison with a pre-specified reference point.	Comparison with measures taken at other points in time on the same person.

of an underlying and fairly static hypothetical construct. This major difference results in the users of each type of test responding quite differently to results. When the user of behavioral assessment instruments obtains differences in scores for an individual over time, the conclusion is that the environment has affected his behavior to produce that difference. Conversely, when scores remain roughly the same, then the behavior has not been changed by the environment. With psychometric assessment, there is little emphasis on obtaining several measures over a period of time (except when determining reliability). A single measure is usually taken, and this should be an indirect measure of a trait or characteristic of the individual. If variability is found, then it is generally considered to be measurement error, rather than a change in the trait.

Cone and Hawkins (1977) conclude that the contrast between the view of traditional assessment--that behaviors are a product of internal events--and the view that behaviors are a product of current observable events, has immense implications for clinical assessment. They state:

> The implications of these assumptions for clinical assessment are considerable. First, for the traditionalist, observable behavior is a sign of an underlying cause and is of only trivial significance in its own right. A variety of responses might serve as satisfactory indicators of the cause. Second, since the causes are enduring products of early experiences, their behavioral manifestations should be consistent over time. Third, the underlying causes are thought to produce generalized forms of responding. That is, a trait manifests itself in various ways and across a variety of different situations (p. xviii).

This difference in using tests either to measure behavior directly or to measure some underlying variable has led to another major distinction between the two types of assessment. Because psychometric tests are

used as indirect measures of variables that are difficult to determine, and can be manifested by a variety of behaviors, there has been a heavy emphasis of the validation of those instruments. Before a test can be used, there has to be a strong indication that it in fact does measure what it is intended to measure. Therefore, these tests generally have accompanying literature describing the validity research that can be used to determine the usefulness of a test for a specific purpose. Behavioral assessment has put very little emphasis on identifying the validity of its instruments because they are seen as measuring a behavior directly. If a reliable count is made of the exact occurrence of a specifically defined behavior, then the question of whether the assessment measures what it intends to measure has much less importance than with psychometric assessment. As Simeonsson (1977, p. 33) states, the functional assessment of behaviors through direct observation has high face validity. Both types of assessment value reliability indicators, although their procedures for determining reliability are quite different. However, validation of behavioral assessment devices is beginning to get some attention.

Goldfried and Linehan (1977, p. 20) indicate that there are several occasions when psychometric issues should be considered in behavioral assessment. Specifically, these are when observations are:

1. Combined to produce a summary score
2. Used to make inferences about behavior in other situations
3. Interrelated and used to make inferences about functional relationships

However, even with the budding emphasis on validation of behavioral assessment, there has been little research on predictive validity or on standardization.

Livingston (1972; 1977) provides a discussion of how reliability and validity are used by both types of assessment, and how validity measures may improve behavioral assessment. Additionally, Goldfried and Linehan (1977) present procedures for establishing content validity of behavioral measures.

There are deviations from these rather diverse starting points for both traditional and behavioral assessment which make them appear more similar. For instance, behavioral assessment may determine 'anxiety' through various direct behavioral measures. Anxiety is then seen as an underlying variable, although it is still seen to be a function of the environment and therefore alterable through behavioral techniques. On the other hand, some psychometric techniques, e.g., intelligence tests, have been used to evaluate the effects of various intervention strategies, and thus are seen as being partially affected by environmental events. However, there do seem to be some fairly major differences between the two types of assessment, and it is not difficult to classify a given test as being one or the other.

The application of these theoretical distinctions among types of assessment needs to be related to the area of screening and diagnosis of at-risk infants and young children. Cross (1977) defines screening as ". . . a measurement activity which identifies in the general population

whose children that appear to be in need of special services in order to deve'op their maximum potential" (p. 4). Its purpose is to sort out children who require furt'er assessment. Diagnosis is then defined as ". . . looking at a child and his environment in depth for four basic purposes:

1. To determine whether a handicapping condition (or conditions) exists
2. To clarify the causes of the identified problem (i.e., is the child nonverbal due to a hearing impairment, mental retardation, an information-process problem, or lack of verbal stimulation)
3. To develop a treatment plan
4. To ascertain the most appropriate service that the program can render the child" (p. 6-7)

Diagnosis may or may not result in treatment or education prescriptions.

For the purpose of this paper, both screening and diagnosis will be limited to procedures concerned with cognitive and perceptua'/motor functioning. Explicitly excluded are medical, sensory, and neurological tests.

Brief Description of Behavioral Research Program

The research program proposed by the Institute will involve a longitudinal approach:

1. To the development of systematic and accurate procedures for identifying children and infants who are at-risk for handicapping conditions

2. To the measurement of environmental events in home and school situations for the purpose of identifying those variables related to handicapping conditions
3. To the development of and validation of intervention strategies which alleviate handicapping conditions in these children

The first two major areas of research need to include children who range from normal and superior to those who are severely handicapped. A great deal of descriptive information on these children and their environments will be collected so that these variables can be evaluated for their utility in the prediction of potential handicapping conditions. The third major research effort, intervention strategies, will develop procedures that affect either environmental variables or the child's learning strategies. The effect of these procedures on the children's development of and severity of handicapping conditions will be measured.

This research program will concentrate on the measurement of direct, observable behaviors and environmental variables. The section of research concerned with the identification of "at-risk" children will emphasize three major variables:

1. The infant's possession and acquisition of receptive language
2. The older child's acquistion of a complex learning task
3. The analysis of discrete motor responses

The section concerned with the relationship of environmental events to at-risk and handicapping conditions will define and evaluate a large number of environmental events, child behavior, and adult behavior through

direct observation of those environments and individuals. The third section, intervention strategies, will manipulate environmental and learning variables and measure the effect on the child. The emphasis here is on within subject comparison of that subject's responses to the manipulations, although the generality of the intervention across individuals will also be examined.

Within all of these primary research areas, there is an emphasis on the direct observation and measurement of the child's behavior and the surrounding environmental stimuli and events.

Relationship of Traditional Assessment to This Behavioral Research Program

Although the emphasis of this Research Institute is on behavioral assessment, there is a need for an auxiliary program of traditional or psychometric assessment to accompany the behavioral assessment and intervention strategies developed by the Institute.

Simeonsson (1977) differentiates three types of early childhood assessment: achievement tests, functional tests, and competency tests. Achievement tests (p. 28-32) are essentially traditional, standardized tests which reflect the achievement of a child relative to his age norms. They are standardized, widely used, and therefore capable of communicating among many test users. However, since their standardization procedures and the resulting norms typically have not included handicapped or at-risk children, they have limited applicability for use as a program

basis. They rely on typical rather than optimal performance, which limits their utility in describing total capabilities of children. Additionally, they have little use in developing educational programs.

Functional assessment (Simeonsson, 1977, pp. 32-34) involves the direct observation of a child interacting with the environment. It seeks to determine the functional relationships of the child's behavior with antecedent, contiguous, and consequent events. It can be tailored to an individual child and then events can be manipulated to determine the causal relationships which may obtain. This same flexibility in adapting to individual situations limits its generality and, in some respects, the ability to communicate to assessment users. It is also demanding of the assessor's time. Competency assessment (Simeonsson, 1977, pp. 34-37) assumes a fixed developmental sequence against which a child can be compared. This approach focuses on the individual, but also gives comparative information for communication with other individuals. It lends itself readily to prescription of educational programs by identifying existing skills, and indicating what the next steps in the developmental sequence are. It assumes a specific theoretical perspective, and can also be time-consuming since it emphasizes optimal rather than typical performance.

The emphasis of the Institute is upon the development of behavioral diagnostic tests which are functional, and in most cases, competency-oriented. Each assessment procedures will be developed to recommend individualized prescriptive programs. The relationship of this information to traditional screening and diagnostic tests will be determined where possible.

This program of traditional assessment is to be undertaken for the following reasons:

1. The availability of standardized test information for children used subjects in the research program will provide a common method for communicating to others what type of subjects they are (Evans & Nelson, 1977, p. 369). Although it is still possible to use data generated through the research effort to describe these children, this will not provide information on how the children compare to others within the normal population in a systematic way. This information will be quite important to many potential consumers of our research efforts, and can be obtained through the use of a normative, standardized test battery.
2. A standardized test battery given periodically to children within the research population can be used as a partial check of the effectiveness of the intervention procedures developed (Evans & Nelson, 1977, pp. 639 and 644) and as a validity measure of the assessment devices developed. That is, if children demonstrate increases in scores on the test battery concurrently with the application of some intervention procedures, then this will add strength to the conclusion that the intervention procedure alleviated to some extent the at-risk or handicapping condition. Additionally, the relationship between assessment procedures developed by the Institute and the results of the traditional test battery will have to be examined to see if the information

is congruent. This will result in correlational information between the traditional tests and the variables examined by the Institute. It must be realized, however, that information obtained this way can only provide added strength or cast doubt about the conclusions drawn through the research effort. It can not be used for determining causal relationships.

3. Another benefit of the availability of the results of periodic testing of children within the research sample is that this information can be included in the data base management system. This information may then be quite useful for researchers at other institutes who may want access to our data base. More important, this information will be used to analyze relationships among variables within the data base to see what variables may warrant more systematic research efforts. It may provide information on the impact of an intervention program on variables or test items other than those directly affected (Evans & Nelson, 1977, p. 644). For instance, if it is found that children who respond one way on a specific test item seem to come from homes within which parents respond to their child in a certain way, then this relationship may be investigated systematically. It is anticipated that the availability of this data base management system will be a potential source of interesting research questions.
4. A final benefit of the traditional test information is that it is possible that it may be useful as a screening procedure for

determining which children, especially those outside of the classrooms operated by investigators within the Institute, may be most appropriate for further prescriptive testing. By eliminating children who are at the extreme ends of the distribution of scores for the assessment devices, only those children who are more likely to benefit from further assessment and intervention will receive it. That is, children who score within one standard deviation either way from the mean score for that age can undergo further assessment. Those who score quite low are those who are already identified as handicapped, and can be included for testing where appropriate for the investigator's goals. Those who score quite high will be excluded since the probability of their being at risk is quite low.

In conclusion, the use of a traditional assessment battery will provide valuable information to the researchers within the Institute, and to persons interested in the research and products developed by the Institute. For this reason, a systematic effort will be made to include a periodic assessment of every research subject within the Institute. Details on the procedures to be used in this effort can be found in the Procedures section.

Reference Notes

1. Leland, H., Nihira, K., Foster, R., Shellhaas, M., & Kagin, E. Conference on Measurement of Adaptive Behavior: II. (Monograph published at Parsons State Hospital and Training Center, Parsons, Kansas, 1966).

2. Foster, R. Evaluation of the mentally retarded. Paper presented at the International Seminar on Mental Retardation, Madison, WI, July 1972.

References

Ciminero, A. R. Behavioral assessment: An overview. In A. R. Ciminero, K. S. Calhoun, & H. E. Adams (Eds). Handbook of behavioral assessment. New York: John Wiley and Sons, 1977.

Cone, J. D., & Hawkins, R. P. Behavioral assessment: New directions in clinical psychology. New York: Brunner/Mazel, 1977.

Cronbach, L. J. Essentials of psychological testing (3rd ed.). New York: Harper and Row Publishers, 1970.

Cross, L. An introduction. In L. Cross, & K. Goin (Eds), Identifying handicapped children: A guide to casefinding, screening, diagnosis, assessment and evaluation. New York: Walker and Company, 1977.

Evans, I. M., & Nelson, R. O. Assessment of child behavior problems. In A. R. Ciminero, Calhoun, K. S., & H. E. Adams (Eds.), Handbook of behavioral assessment. New York: John Wiley and Sons, 1977.

Goldfried, M. R., & Linehan, M. M. Basic issues in behavioral assessment. In A. R. Ciminero, K. S. Calhoun, & H. E. Adams (Eds.), Handbook of behavioral assessment. New York: John Wiley and Sons, 1977.

Hammill, M. R., & Larsen, S. C. The effectiveness of psycholinguistic training. Exceptional Children, 1974, 41, 5-15.

Hammill, D. D., & Wiederholt, J. L. Review of the Frostig Visual Perceptual test and the related training program. In L. Mann & D. Sabatino (Eds.), The first review of special education. Philadelphia: Journal of Special Education Press, 1973, pp. 33-48.

Livingston, S. A. Criterion-referenced applications of classical test theory. Journal of Educational Measurement, 1972, 9(1), 13-26.

Livingston, S. A. Psychometric techniques for criterion-referenced testing and behavioral assessment. In J. D. Cone, & R. P. Hawkins (Eds.), Behavioral assessment: New directions in clinical psychology. New York: Brunner/Mazel, 1977.

Sedlak, R. A., & Weener, P. Review of research on the Illinois Test of Psycholinguistic Abilities. In L. Mann & D. A. Sabatino (Eds.), The first review of special education. Philadelphia: Journal of Special Education Press, 1973, pp. 113-164.

Simeonsson, R. J. Infant assessment and developmental handicaps. In B. M. Caldwell, & Stedman, D. J. (Eds.), Infant education: A guide for helping handicapped children in the first three years. New York: Walker and Company, 1977.

Sulzer-Azaroff, B., & Mayer, G. R. Applying behavior-analysis procedures with children and youth. New York: Holt, Rinehart and Winston, 1977.

Tillman, M., Bersoff, D., & Dolly, J. Learning to teach: A decision-making system. Lexington, MA: D. C. Heath and Co., 1976.

Ysseldyke, J. E. Diagnostic-prescriptive teaching: The search for aptitude-treatment interaction. In L. Mann, & D. A. Sabatino (Eds.), The first review of special education. Philadelphia: Journal of Special Education Press, 1973, pp. 5-33.